

Biased diffusion on Japanese inter-firm trading network: Estimation of sales from network structure

Hayafumi Watanabe¹, Hideki Takayasu² and Misako Takayasu¹

¹Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, 4529 Nagatsuta-cho, Midori-ku, Yokohama 226-8502, Japan

² Sony Computer Science Laboratories, 3-14-13 Higashi-Gotanda, Shinagawa-ku, Tokyo 141-0022, Japan

E-mail: h-watanabe@smp.dis.titech.ac.jp

Abstract. To investigate the actual phenomena of transport on a complex network, we analysed empirical data for an inter-firm trading network, which consists of about one million Japanese firms and the sales of these firms (a sale corresponds to the total in-flow into a node). First, we analysed the relationships between sales and sales of nearest neighbourhoods from which we obtain a simple linear relationship between sales and the weighted sum of sales of nearest neighbourhoods (i.e., customers). In addition, we introduce a simple money transport model that is coherent with this empirical observation. In this model, a firm (i.e., customer) distributes money to its out-edges (suppliers) proportionally to the in-degree of destinations. From intensive numerical simulations, we find that the steady flows derived from these models can approximately reproduce the distribution of sales of actual firms. The sales of individual firms deduced from the money-transport model are shown to be proportional, on an average, to the real sales.

PACS numbers: 89.75.Hc, 89.75.Da, 05.10.Gg

1. Introduction

The circulation of money is often likened to the circulation of blood. For instance, in the middle of the 18th century, French physiocrat François Quesnay introduced his economic theory ‘*tableau économique*’, which is one of the important foundations of modern economics, from the theory of blood circulation described by William Harvey in the 17th century [1]. If this analogy is acceptable, what are the differences between the flow of money (i.e., the flow within society) and the flow of blood (i.e., the flow within the body) in terms of transport phenomena?

Transport processes such as diffusion, advection and radiation, play a fundamental role in physics and theories of transport processes are widely applied in chemistry, biology, engineering, etc. One transport problem that has long attracted interest is transport in complex systems such as biological or the social systems. Because of recent accumulations of data regarding complex systems and developments in the theory of complex networks, we can now provide a new perspective on such problems.

Complex networks have been studied intensively over the past decade [2]-[4]. These studies revealed that complex networks can be observed in a wide range of real systems both in natural and man-made. In particular, transport phenomena in complex networks have been investigated. For example, random walks on complex networks have been studied from various viewpoints [5, 6, 9]. The PageRank, which is one of the most successful indices evaluating the importance of web pages and which is applied by Internet search engines, corresponds to the steady-state density of transport caused by random walks on the World Wide Web. Other types of transportation on complex networks have also been studied [7, 8, 10].

What properties characterize the actual transport on a complex network? A great majority of studies of transportation on complex networks has been based on theoretical approaches; however, a few studies have been involved actual transportation on complex networks. For example, R. Guimera *et al.* revealed the nonlinear relationships between degrees of airports traffics on the world-wide-airport network [11], P. Sen *et al.* studied the number of trains on the Indian rail way network [12] and A. Chmiel *et al.* investigate on visitors on portal sites and self-attracting walks can well describe its properties [13].

To investigate real phenomena of transport on a complex network, we focus on the transport of money on the trading network where firms correspond to network nodes and the trade relations between firms correspond to network edges. This system is useful for studying the real transport phenomena on the complex network, because we can estimate the total in-flow of each node by from the sales data.

In this study, we show that we can consistently estimate the sales from the structure of the inter-firm trading network, which is analogous to estimating blood flows from the vascular structure. In sec. II, we start by analysing the trading network data, from about 900,000 Japanese firms and their corresponding sales data. Next, we introduce two transport models and discuss their properties in Sec. III. In Sec. IV, we compare the steady-state sales of the model with the actual sales and show that flows generated

by the models can reproduce the well-known Zipf's law; namely, that the cumulative distribution of sales obeys a power-law distribution with the exponent close to -1. Finally, we conclude with a discussion in Sec. V.

2. Data analysis

The data set was provided by Tokyo Shoko Research, Ltd. (TSR). and contains about one million firms practically covering all active firms in Japan. For each firm, the data set contains the annual sales and a list of business partners, categorized into suppliers and customers [14, 15].

From this list, we generated a network (firm network), whose nodes are the firms and the edges are defined by the following rule: If the i -th firm buys something from the j -th firm, or equivalently if money flows the i -th to the j -th, we connect from the i -th to the j -th with a directed link[14]. Figure 1(a) shows the average degree of the nearest neighbours, which is denoted by a function of degree k : $k_{nn}(k)$ [16]. The data shown in this figure confirms that the firm network has a negative degree-degree correlation.

To clarify the properties of this firm network, we perform a parallel analysis using an artificial random network having the same degree distribution as the firm network [14, 17, 18]. We generate an artificial random network by using the Markov-chain Monte Carlo switching algorithm[19] which is used to repeatedly choose the edge pairs randomly and switch from ' $X1 \rightarrow Y1$ and $X2 \rightarrow Y2$ ' to ' $X1 \rightarrow Y2$ and $X2 \rightarrow Y1$ ' until the network is well randomized. This artificial network, which we call the shuffled network in this paper, is an almost uncorrelated network having the same degree distribution as the original firm network. The red line in figure 1(a) shows the degree-degree correlation of this network. By comparing the differences in behaviour between the real firm network and the shuffled network, we can check the effect of a correlation.

2.1. Statistical properties of individual firms

We begin by investigating the properties of individual firms. figure 1 (b) shows cumulative distributions s , $k^{(in)}$ and $k^{(out)}$, where s is the annual sales in 2005, $k^{(in)}$ is the in-degree and $k^{(out)}$ is the out-degree. The data shown in the figure indicates that s , $k^{(in)}$ and $k^{(out)}$ obey the following power-law cumulative distribution functions (CDFs):

$$P(> s) \propto s^{-\alpha_s} \tag{1}$$

$$P(> k^{(in)}) \propto (k^{(in)})^{-\alpha_{in}} \tag{2}$$

$$P(> k^{(out)}) \propto (k^{(out)})^{-\alpha_{out}} \tag{3}$$

with exponents $\alpha_s = 1.0$, $\alpha_{in} = 1.3$ and $\alpha_{out} = 1.3$. For sales s , this empirical fact, which is well known as Zipf's law, is observed in various countries [20, 21, 22, 14].

Next, we investigate the correlation between sales and degrees. We calculate the conditional mean of s as a function of $k^{(in)}$. As shown in figure 1 (c), for a large in-degree $k^{(in)}$, $\langle s \rangle_{k^{(in)}}$ can be described as a power law

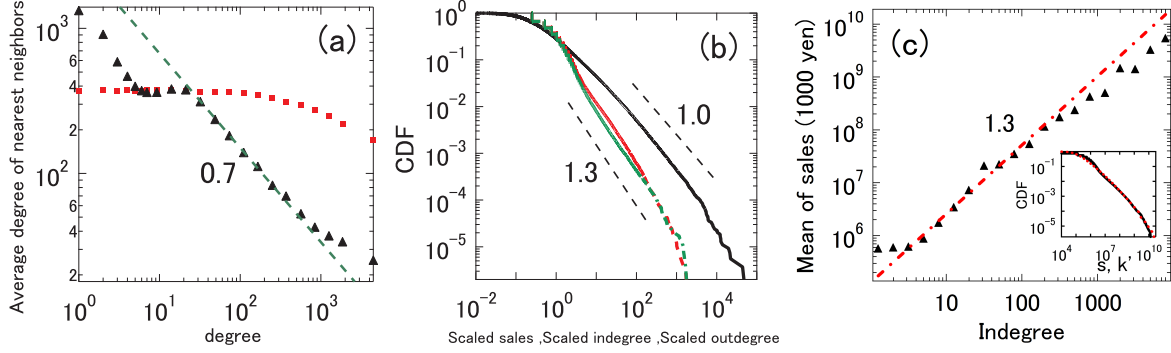


Figure 1. (a) Average degree of nearest neighbours $k_{nn}(k)$ for the LSCC of the firm network (black triangles) and that for the shuffled network (red squares). The LSCC of the firm network has a stronger negative degree-degree correlation than the shuffled network.

(b) The Cumulative distribution of scaled sales (black solid line), scaled in-degrees (red broken line) and scaled out-degrees (green dash-dotted line). Each value is scaled (i.e., divided) by the inter quartile range. The exponent for sales is different from that for in-degrees and out-degrees.

(c) The conditional mean of sales given in-degree (black triangles). The red dash-dot line is $s = 10^{5.1} \cdot k^{(in)1.3}$. The conditional mean is proportional to the 1.3 power of in-degrees. The bottom right figure shows CDFs of s (black solid line) and $k' = 10^{5.1} \cdot k^{(in)1.3}$ (red broken line). These CDFs are in good agreement.

$$\langle s \rangle_{k^{(in)}} \propto k^{(in)\beta_{s|k}} \quad (4)$$

with $\beta_{s|k} \approx 1.3$, where $\langle s \rangle_{k^{(in)}}$ is the conditional mean of s for given $k^{(in)}$. These results imply that the mean of 'sales per in-degree' increases with increasing in-degree.

Roughly speaking, we can theoretically derive a relationship among α_{in} , α_s , $\beta_{s|k}$ as a result of transformation of random variables. Assuming that, $k^{(in)}$ obeys the power-law distribution with the probability density function (PDF)

$$p_{k^{(in)}}(k^{(in)}) \propto k^{(in)-\alpha_{in}-1} \quad (5)$$

, and s and $k^{(in)}$ satisfy the power law relationship

$$s \propto k^{(in)\beta_{s|k}}, \quad (6)$$

then by changing the variables, the PDF of s becomes

$$p_s(s) \propto p_{k^{(in)}}(k^{(in)}) \cdot \left| \frac{dk^{(in)}}{ds} \right| \propto s^{1/\beta_{s|k}(-\alpha_{in}-1)} \cdot s^{1/\beta_{s|k}-1} \propto s^{-\alpha_{in}/\beta_{s|k}-1}. \quad (7)$$

Thus, we get the following nontrivial relationship between power law indices:

$$\beta_{s|k} = \alpha_{in}/\alpha_s. \quad (8)$$

This relationship is consistent with the observed values.

2.2. Nearest-neighbourhood correlations

We now consider the relationships between the sales of a firm and the sales of its customers. Customers of m -th node are defined by the nodes whose out-going edges reach the i -th node. We introduce two kinds of weighted sums of customer sales, $s_m^{(1)}$ and $s_m^{(2)}$:

$$s_m^{(1)} \equiv \sum_{i=1}^N A_{im} \frac{1}{\sum_{j=1}^N A_{ij}} s_i = \sum_{i=1}^N A_{im} \frac{s_i}{k_i^{(out)}} \quad (9)$$

$$s_m^{(2)} \equiv \sum_{i=1}^N A_{im} \frac{\sum_{l=1}^N A_{lm}}{\sum_{j=1}^N A_{ij} \sum_{l=1}^N A_{lj}} s_i = \sum_{i=1}^N A_{im} \frac{k_m^{(in)}}{\sum_{j=1}^N A_{ij} k_j^{(in)}} s_i \quad (10)$$

where, A is an $N \times N$ adjacency matrix defined by

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge from } i \text{ to } j \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

In the case of $s^{(1)}$, a customer distributes money among all its suppliers evenly. However, in the case of $s^{(2)}$, a customer distributes money among its suppliers in proportion to suppliers' in-degree. We calculate conditional mean of $s^{(1)}$ given s , denoted by $\langle s^{(1)} \rangle_s$, and conditional mean of $s^{(2)}$ given s , denoted by $\langle s^{(2)} \rangle_s$, functions of s :

$$\begin{aligned} \langle s^{(1)} \rangle_s &= \frac{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} s_l^{(1)}}{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} 1} \\ &= \frac{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} \left\{ \sum_{i=1}^N A_{il} \frac{s_i}{k_i^{(out)}} \right\}}{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} 1}, \end{aligned} \quad (12)$$

$$\begin{aligned} \langle s^{(2)} \rangle_s &= \frac{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} s_l^{(2)}}{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} 1} \\ &= \frac{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} \left\{ \sum_{i=1}^N A_{il} \frac{k_l^{(in)}}{\sum_{j=1}^N A_{ij} k_j^{(in)}} s_i \right\}}{\sum_{l \in \{l | s - \delta_s^{(b)} \leq s_l < s + \delta_s^{(u)}\}} 1}, \end{aligned} \quad (13)$$

where, $\delta_s^{(b)}, \delta_s^{(u)} \ll s$. In this paper, we estimate other conditional means $\langle \cdot \rangle_s$ in a similar manner.

As shown in figures 2(a) and 2(b), for large sales s , $\langle s^{(1)} \rangle_s$ and $\langle s^{(2)} \rangle_s$ can be described as power laws:

$$\langle s^{(1)} \rangle_s \propto s^{0.8} \quad (14)$$

$$\langle s^{(2)} \rangle_s \propto s^{1.0}. \quad (15)$$

For $s^{(2)}$, in particular, we observe a simple linear relationship in the region above 10^9 yen. Its proportionality constant is equals to about 1.

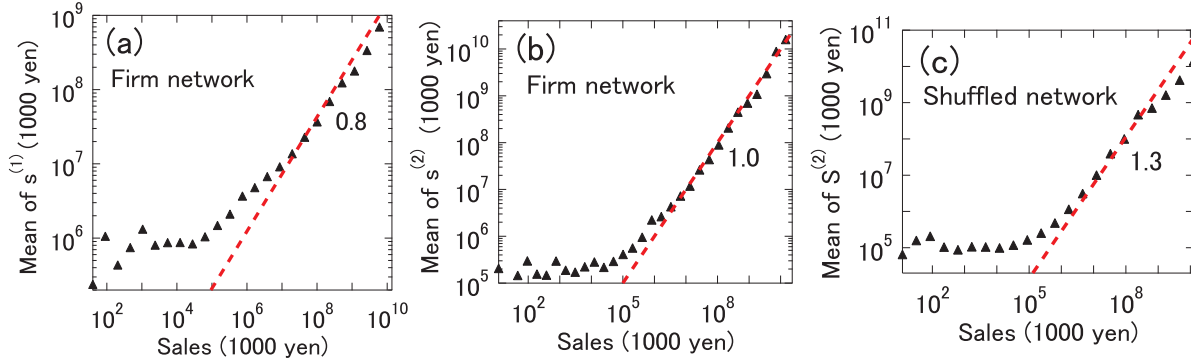


Figure 2. (a)-(c) Correlations between sales and $s^{(i)}$ ($i = 1, 2$). (a) $\langle s^{(1)} \rangle_s$, which is the conditional mean of $s^{(1)}$ given by the sales s , as a function of the sales s for the firm network. Red broken line shows $s^{(1)} = 30 \cdot s^{0.8}$. (b) $\langle s^{(2)} \rangle_s$, which is the conditional mean of $s^{(2)}$ given by the sales s , as a function of the sales s for the firm network. Red broken line shows $s^{(2)} = s$. (c) $\langle s^{(2)} \rangle_s$, which is the conditional mean of $s^{(2)}$ given by the sales s , as a function of the sales s for the shuffled network. Red broken line shows $s^{(2)} = 0.0045 \cdot s^{1.3}$. We confirm that for the firm network, $\langle s^{(2)} \rangle_s$ is almost equal to s for the large sales region. In contrast, $\langle s^{(1)} \rangle_s$ for the firm network and $\langle s^{(2)} \rangle_s$ for the shuffled network are not proportional to s .

The results shown in figure 2 (c) confirm that $\langle s^{(2)} \rangle_s$ is not proportional to s for the shuffled network. In other words, equation (15) does not hold for the shuffled network, which has the same degree distribution as the real firm network.

3. Simulations of transport

3.1. Models

In correspondence with the local relationships given in equations (9) and (10), we introduce the following two models of the time evolution of locally conserved scalar quantities, $x_m(t)$.

Model-1 (PageRank model)

$$x_m(t+1) = \sum_{i=1}^N A_{im} \frac{1}{k_i^{(out)}} x_i(t), \quad (16)$$

Model-2 (Biased distribution model)

$$x_m(t+1) = \sum_{i=1}^N A_{im} \frac{k_m^{(in)}}{\sum_{j=1}^N A_{ij} k_j^{(in)}} x_i(t). \quad (17)$$

Note that for $k_i^{(out)} = 0$ in equation (16) or $\sum_{j=1}^N A_{ij} k_j^{(in)} = 0$ in equation (17), we omit the contributions of the i -th node.

In equation (16), a node(customer) evenly distributes its scalar(money) among its outgoing edges (suppliers); i.e., Model-1 corresponds to the PageRank model [5]. However, Model-2 corresponds to a kind of biased random-walk model [23]. In equation

(17), a node(customer) distributes its scalar(money) to its outgoing edges in proportion to the node in-degrees indicated by the outgoing edges (suppliers).

We apply these time-evolution models to two types of networks. The first network is the largest strongly connected component(LSCC) of the real firm network. A strongly connected component(SCC) is defined as the maximal subset of edges in a network such that each node can reach all others and is itself reachable from all others along a directed path[24]. The LSCC of the firm network is defined as the SCC having the largest number of nodes when we decompose the firm network into SCCs. The LSCC of the firm network is the core of the firm network, containing 462,602 nodes and 2,583,620 edges. In addition, the previously mentioned statistical properties of the firm network holds true for the LSCC of the firm network. Note that for the both Model-1 and Model-2, there is no need to consider a boundary-condition or end effects, because of an SCC does not have exits(nodes having no out-edges) or entrances(nodes having no in-edges). Therefore, we can focus on the properties of models for the bulk of the network, which is why we do not use the original firm network but the LSCC of the original firm network for the first step of the numerical simulations.

The second network is the shuffled LSCC of the firm network, generated by the above-mentioned Markov-chain Monte Carlo switching algorithm[19]. The shuffled LSCC of the firm network is an almost uncorrelated network with the same degree distribution as the LSCC of the firm network. For our simulation, we used the shuffled LSCC of a network that consisted of a single SCC and we also assumed that the network consisted of a single SCC for the theoretical analysis.

3.2. Properties of models

For a strongly connected network, we consider the existence of a steady state for the Model-1 and Model-2. Because SCCs have no outlets, the total scalar $\sum_{i=1}^N x_i(t) = \sum_{i=1}^N x_i(0)$ is conserved in both models. For the steady state with normalization, $\sum_{i=1}^N x_i(0) = \sum_{i=1}^N x_i(t) = 1$, both models are Markov-chains, where the probability of existence of the m -th node, $p_m(t)$ is given by $x_m(t)$ and the transition probability Q_{mi} from the i -th node to the m -th node is $A_{im} \cdot 1/k_i^{(out)}$ and $A_{im} \cdot k_m^{(in)} / \sum_{j=1}^N (A_{ij} k_j^{(in)})$ for the Model-1 and Model-2, respectively. These Markov-chains are irreducible; i.e., there is a non-zero transition probability from any state to any other state. This property arises because a path exists between any two nodes in the graph on the strongly connected network and the transition probability from the i -th node to the j -th node is a non-zero for the node pair the i -th and the j -th such that $A_{ij} = 1$. In general, it is known that an irreducible Markov-chain with the finite number of states have unique steady state [25], therefore, both models have the unique steady state for the strongly connected network. We denote this steady state for the given strongly connected network by $p^{(s)}$. According to linearity, the steady state $x^{(s)}$ is obtained as

$$x_m^{(s)} = \left(\sum_{i=1}^N x_i(0) \right) \cdot p_m^{(s)} \quad (m = 1, 2, \dots, N). \quad (18)$$

where $\sum_{i=1}^N x_i(0)$ is the total sum of the initial values.

3.2.1. Model-1 (PageRank model) To understand the properties of our model for the LSCC of the firm network and the shuffled LSCC of the firm network, we simulate the Model-1 by equation (16). Starting with the initial condition $x_i(0) = 1$ ($i = 1, 2, \dots, N$), the CDF of x converges to the steady state distribution as shown in figure 3(a). For both cases the real and shuffled network, the distribution of x follows a power law with an exponent of about 1.3, which is the power-law exponent for in-degree α_{in} . Thus, Model-1 is not consistent with the empirical sales distribution because the empirical sales distribution follows the Zipf's law with exponent -1. Figure 3(b) shows the conditional mean of x for given $k^{(in)}$ as a function of $k^{(in)}$, which is denoted by $\langle x \rangle_{k^{(in)}}$. For both networks, we get the following liner function:

$$\langle x \rangle_{k^{(in)}} \propto k^{(in)}. \quad (19)$$

This result can be explained by the mean-field solution of PageRank [9]. However, it is inadequate to regard x as sales because equation (19) disagrees with the empirical relationship, equation (4).

3.2.2. Model-2 (Biased diffusion model) We apply the same analysis to Model-2. Starting with the same initial condition as for Model-1, the CDF of x converges quickly to a power law, as shown in figure 3 (c) for both case of the LSCC of the real firm network and for the shuffled LSCC. The exponent of the steady state power law is about -1 for the LSCC of the firm network, which agrees with the empirical observation for sales, equation (1). However, for the shuffled LSCC of the firm network, the exponent takes -0.65, which disagrees with the empirical observation equation (1). Note that the steady state of Model-2 is sensitive to the correlation of the network structure.

figure. 3 (d) shows the conditional mean of x for given $k^{(in)}$ viewed as a function of $k^{(in)}$, denoted by $\langle x \rangle_{k^{(in)}}$. For all cases, its behaviour approximately follows the power law:

$$\langle x \rangle_{k^{(in)}} \propto \{k^{(in)}\}^{\beta_{x|k}}, \quad (20)$$

with exponent $\beta_{x|k}$ satisfying equation (8). For the shuffled LSCC of the firm network, the exponent $\beta_{x|k}$, which takes a value about 2, is explained by an annealing approximation solution of the biased random walk model for an uncorrelated network [23]. However, this exponent disagrees with the empirical exponent given in equation (4). Conversely, for the LSCC of the firm network, the exponent $\beta_{x|k}$, which is equal to 1.3, agrees with empirical exponent given in equation (4). These results indicate that the properties of the power-law exponent are connected to the space-correlation of the network.

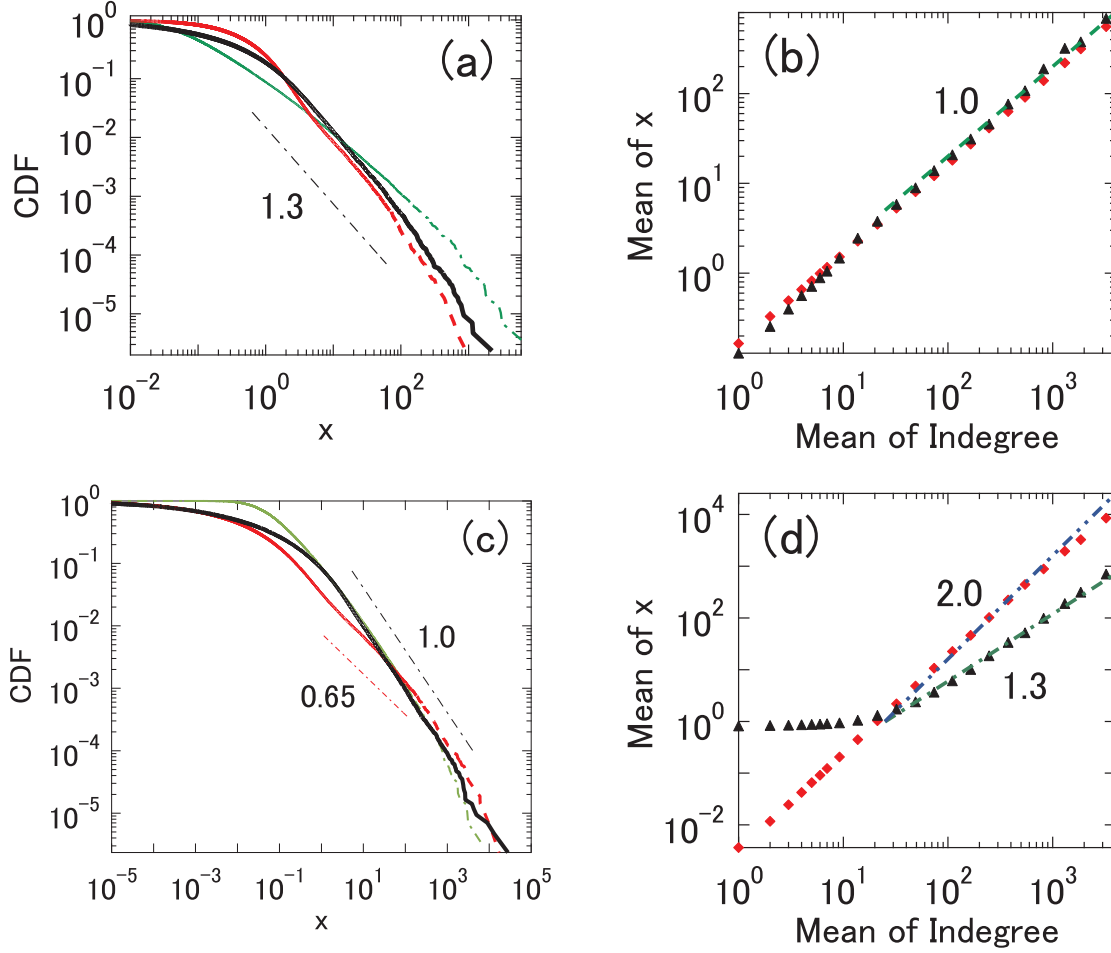


Figure 3. (a) CDFs of x in the case of Model-1 for the large time t ($t = 2^{15}$). Shown are for the LSCC of the firm network (black solid line) and for the shuffled LSCC of the firm network (red dashed line). Sales of real firms scaled by the sum of sales are shown the green dash-dotted line. For both networks, x obeys a power law distribution with an exponent about 1.3.

(b) The conditional mean of x given the in-degree $k^{(in)}$, denoted by $\langle x \rangle_{k^{(in)}}$ in the case of Model-1. Data shown are for the LSCC of firm network (black triangle), for the shuffled LSCC of the firm network (red diamond) and $x \propto k^{(in)}$ (broken green line).

In all cases, the conditional mean is proportional to in-degrees (i.e., a power-law exponent of 1). (c) The CDFs of x for Model-2 for the large time t ($t = 2^{15}$). Data shown are for the LSCC of the firm network (black solid line) and for the shuffled LSCC of the firm network. Sales of real firms scaled by the sum of sales are shown by the green dash-dotted line. For the firm networks, x obeys the power law distribution with exponent about 1. The shuffled networks has with an exponent of 0.65, which corresponds to the exponent of the annealing approximation of the biased random walk model for the uncorrelated scale free network with the exponent 1.3. The CDF of x agrees with the actual scaled sales for the large s , only for the firm network.

(d) The conditional mean of x given in-degree $k^{(in)}$, denoted by $\langle x \rangle_{k^{(in)}}$ for Model-1. Data shown are for the LSCC of the firm network (black triangles), for the shuffled LSCC of the firm network (red diamonds), $x \propto k^{(in)1.3}$ (green dash-dotted line) and $x \propto k^{(in)2}$ (blue dash-double-dotted line).

4. Comparisons between simulations and observations

In this section, we consider the statistical properties of the entire firm network. For general networks, the steady states of the Model-1 and Model-2 do not always exist. In addition, because the scalar quantity defined for the nodes flows into nodes that do not have out-edges, most nodes in the network bulk have zero scalar after many time steps. To obtain a non-trivial steady state we add the effects of injection and dissipation in the following forms:

Model-1 (PageRank model)

$$x_m(t+1) = r \sum_{i=1}^N A_{im} \frac{1}{k_i^{(out)}} x_i(t) + f, \quad (21)$$

Model-2 (Biased diffusion model)

$$x_m(t+1) = r \sum_{i=1}^N A_{im} \frac{k_m^{(in)}}{\sum_{j=1}^N A_{ij} k_j^{(in)}} x_i(t) + f \quad (22)$$

where $0 < 1 - r \leq 1$ is the dissipation factor and $f > 0$ is the injection term, which is a constant and positive value. Note that, for $k_i^{(out)} = 0$ in equation (21) or $\sum_{j=1}^N A_{ij} k_j^{(in)} = 0$ in equation (22), we omit contributions of the i -th node.

In general, starting from any initial state, the time evolution given by $x(t+1) = rBx(t) + f$ converges to a unique steady state provided that the maximum eigen-value of rB less than 1 [26], where $x(t)$ is a state vector for time t and rB is a square matrix. Denoting the maximum eigen-value of rB as λ and the corresponding eigen-vector as y gives

$$|\lambda| \sum_{i=1}^N |y_i| = \sum_{i=1}^N |(rBy)_i| = r \sum_{i=1}^N \left| \sum_{j=1}^N B_{ij} y_j \right| \quad (23)$$

$$\leq r \sum_{i=1}^N \sum_{j=1}^N |B_{ji} y_i| = r \sum_{i=1}^N |y_i| \sum_{j=1}^N |B_{ji}| \quad (24)$$

From this definition, we obtain the following for the Model-1

$$\sum_{j=1}^N |B_{ji}| = \begin{cases} \sum_{j=1}^N A_{ij} \frac{1}{k_i^{(out)}} = \frac{k_i^{(out)}}{k_i^{(out)}} = 1 & (k_i^{(out)} \neq 0) \\ 0 & (k_i^{(out)} = 0) \end{cases} \quad (25)$$

and for Model-2, we obtain

$$\sum_{j=1}^N |B_{ji}| = \begin{cases} \sum_{j=1}^N A_{ij} \frac{k_j^{(in)}}{\sum_{l=1}^N A_{il} k_l^{(in)}} = \frac{\sum_{j=1}^N A_{ij} k_j^{(in)}}{\sum_{l=1}^N A_{il} k_l^{(in)}} = 1 & (k_i^{(out)} \neq 0) \\ 0 & (k_i^{(out)} = 0) \end{cases} \quad (26)$$

Thus, in both cases,

$$r \sum_{j=1}^N |B_{ji}| < \sum_{j=1}^N |B_{ji}| \leq 1 \quad (27)$$

Substituting equation (27) in equation (24), we obtain

$$|\lambda| \sum_{i=1}^N |y_i| \leq \sum_{i=1}^N |y_i| \sum_{j=1}^N |r B_{ji}| < \sum_{i=1}^N |y_i| \quad (28)$$

Thus,

$$|\lambda| < 1 \quad (29)$$

Therefore, starting from any initial state, x converges to a unique steady state.

In figure 4 we compare the CDFs between the simulations and observation. Figure 4(a) shows the results for case of Model-1 for the firm network, figure 4(b) shows a results for the case of Model-2 for the firm network and figure 4(c) shows the results for Model-2 for the shuffled network. For these figures, we used $r = 0.95$ and $f = 1.33 \cdot 10^5$ (1000 yen). Under these conditions, $\sum_{i=1}^N s_i = \sum_{i=1}^N x_i(\infty)$ holds. From figure 4(a) and (b), we see that the CDF of x for Model-2 agrees well with the sales distribution observed for the firm network, whereas, the CDF of x for Model-1 disagrees with the observed CDF. In addition, the result shown in figure 4(c), implies that Zipf's law, which is observed in actual data, is strongly related to the degree-degree correlation mentioned in the preceding section.

Next, we compare the results for x from our simulation with the sales of the real firms one by one. figure 4(d), 4(e) and 4(f) show the actual firm sales on the vertical axis and the result for x obtained from the above-mentioned network-flow on simulation in the horizontal axis. Figure 4(d) shows the result for Model-1 for the firm network, figure 4(e) shows the same for Model-2 for the firm network and figure 4(f) shows the result for Model-2 for the shuffled network. Figure 4(e) show that for large x , in the case of Model-2, the conditional mean of actual sales s given sales of the simulation x as a function of x , denoted by $\langle s \rangle_x$ is almost equal to x . Meanwhile, figure 4(d) shows that for Model-1 applied to the the firm network, $\langle s \rangle_x$ is proportional to about the 1.3 power of x , whereas for Model-2 applied to the shuffled network, proportional to about the 0.7 power of x , as shown in figure 4(f). In both cases, $\langle s \rangle_x$ is not proportional to x . This result suggests that Model-2(i.e., the model with injection and dissipation) applied to the firm network roughly reproduces the values of sales of actual firms for simulation sales x larger than about $3 \cdot 10^6$ (1000 yen).

5. Conclusion and discussion

In this paper, we demonstrated that we can roughly estimate the sales of firms from the structure of the Japanese inter-firm trading network. First we found the simple linear local relationship between sales of a firm and the weighted sum of sales of its customers by analysing data from the Japanese inter-firm trading network and corresponding sales

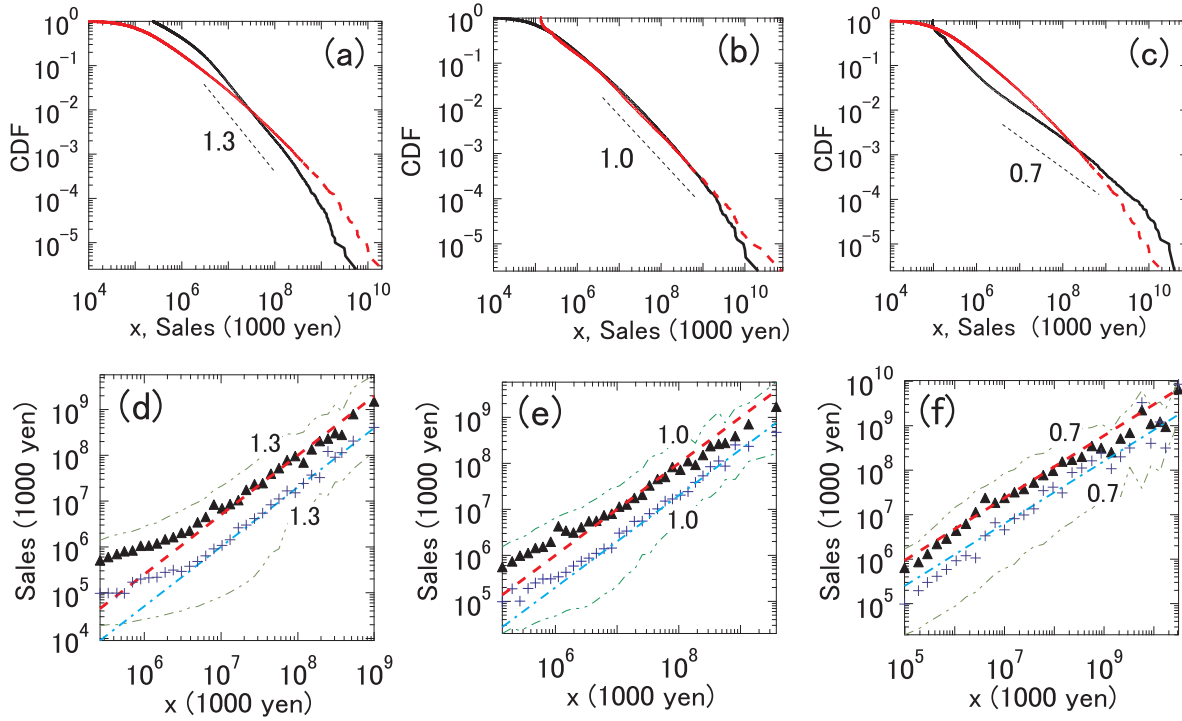


Figure 4. (a)-(c) Comparisons of the cumulative distribution between simulations and observations of actual sales. We use the value $r = 0.95$ and $f = 1.33 \cdot 10^5$ (1000 yen). Data shown are x (solid black line) and actual sales (broken red line). (a) Model-1 for the firm network, (b) Model-2 for the firm network and (c) Model-2 for the shuffled network. For (a) and (c), the CDF of the sales s disagrees with the CDF of simulation x ; however, for (b), CDFs of the sales s agrees with x .

(d)-(f) Correlation between x and sales of actual firms s . Data shown are the conditional mean of sales s given x , denoted by $\langle s \rangle_x$ (black triangles), the conditional mode of sales given x (blue crosses) and the conditional 5 percentile and the 95 percentile (green dashed-double-dotted line). (d) Model-1 for the firm network (broken red line: $s = 0.004 \cdot x^{1.3}$, blue dash-dot line: $s = 0.0008 \cdot x^{1.3}$), (e) Model-2 for the firm network (red broken line: $s = x$, blue dash-dotted line: $s = 0.2 \cdot x$) and (f) Model-2 for the shuffled network (red broken line: $s = 300 \cdot x^{0.7}$, blue dash-dotted line: $s = 80 \cdot x^{0.7}$).

data. Next, we introduced a model (Model-2) that satisfies this local linear relationship between adjacent nodes. In this model, a firm (customer) distributes money to its out-edges (suppliers) proportionally to the in-degree of destinations. By using this model to numerically simulate the real firm network, we confirmed that the steady flows derived from the money-transport model reproduce the distribution of real firm sales and sales of individual firms on average. In addition, we also confirmed that the PageRank (Model-1), which corresponds to the equal distribution of money to out-edges, does not reproduce the distribution of sales. Note that Model-2 corresponds to a biased random walk whose transition probabilities are proportional to the in-degrees of destinations. Therefore, based on our model, we argue that actual firm sales are proportional, on an average, to the existence probability (or the mean stay time) for the steady state of a biased random walker on the firm network.

Applied to the firm network, Model-2 explains the difference between the exponent of the distribution of in-degrees and that of sales, and it also reproduces Zipf's law for the firm network. However, on application to the shuffled network, which is almost uncorrelated network with the same degree distribution as the real firm network, we also confirmed that sales derived by the model in the steady state does not obey Zipf's law. In other words, in our framework, we need not only a particular transport model but also a particular network structure to reproduce Zipf's law. Thus, considering only the effect of the transport mentioned in this paper is sufficient to explain the universal Zipf's law.

For the Model-2 simulation, the power-law exponent is influenced by the degree-degree correlation, whereas this dependence is not observed for Model-1 (PageRank model). Although many actual complex networks have the degree-degree correlation [4], few studies exist of biased random walk on correlated complex networks. A detail survey of this dependence will be reported in a future presentation.

Acknowledgments

The authors are grateful to the Research Institute of Economy, Trade and industry (RIETI) for allowing us to use the TSR data. This work is partly supported by Grant-in-Aid for JSPS Fellows Grant No. 219685 (H.W.) and Grant-in-Aid for Scientific Research No. 22656025 (M.T.) from JSPS. The numerical calculations were carried out on TSUBAME at Global Scientific Information and Computing Center of Tokyo Institute of Technology.

References

- [1] Johnson J 1966 *J. Finance* **21** 119-22
- [2] Albert R and Barabási A-L 2002 *Rev. Mod. Phys.* **74** 47-97
- [3] Newman M E J 2003 *SIAM Rev.* **45** 167-256
- [4] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D 2006 *Phys. Rep.* **424** 175-308
- [5] Brin S and Page L 1998 *Computer Networks and ISDN Systems* **30** 107-17
- [6] Noh J D and Rieger H 2004 *Phys. Rev. Lett.* **92** 118701
- [7] Tadi B, Rodgers G J and Thurner *IJBC* **17** 2363-85
- [8] Carmi S, Wu Z, Lopez E, Havlin S and Eugene Stanley H 2007 *Eur. Phys. J.B* **57** 165-74
- [9] Aiello W, Broder A, Janssen J, Milos E, Fortunato S, Boguna M, Flammini A and Menczer F 2008 *Algorithms and Models for the Web-Graph (Lecture Notes in Computer Science vol 4936)* ed W Aiello, *et al* . (Berlin Heidelberg: Springer Berlin Heidelberg) p 59-71
- [10] Herrero C 2005 *Phys. Rev. E* **71** 06103
- [11] Guimerà R, Mossa S, Turtschi A and Amaral L A N 2005 *Proc. Natl. Acad. Sci. USA* **102** 7794-99
- [12] Sen P, Dasgupta S, Chatterjee A, Sreeram P, Mukherjee G and Manna S 2003 *Phys. Rev. E* **67** 036106
- [13] Chmiel A, Kowalska K and Holyst J A 2009 *Phys. Rev. E* **80** 066122
- [14] Ohnishi T, Takayasu H and Takayasu M 2009 *Prog. Theor. Phys. Suppl.* **179** 157-66
- [15] Fujiwara Y and Aoyama H 2010 *Eur. Phys. J.B* **77** 565-580
- [16] Pastor-Satorras R Vazquez A and Vespignani A 2001 *Phys. Rev. Lett.* **87** 258701
- [17] Boas P R V, Rodrigues F A, Travieso G and Costa L F 2008 *Phys. Rev. E* **77** 026106

- [18] Doye J P K and Massen C P 2005 *J. Chem. Phys* **122** 84105
- [19] Milo R, Kashtan N, Itzkovitz S, Newman M E J and Alon U 2003 *Preprint* cond-mat/0312028
- [20] Pareto V 1896 *Cours d' Economie Politique* (Geneva: Droz).
- [21] Axtell R L 2001 *Science* **293** 1818-20
- [22] Fujiwara Y, Diguilmi C, Aoyama H, Gallegati M and Souma W 2004 *Physica A* **335** 197-216
- [23] Fronczak A and Fronczak P 2009 *Phys. Rev. E* **80** 016107
- [24] Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A and Wiener J 2000 *Computer Networks* **33** 309-20
- [25] C D Meyer 2000 *Matrix Analysis and Applied Linear Algebra* (Philadelphia: SIAM)
- [26] Saad Y 2003 *Iterative methods for sparse linear systems* (Philadelphia: SIAM)